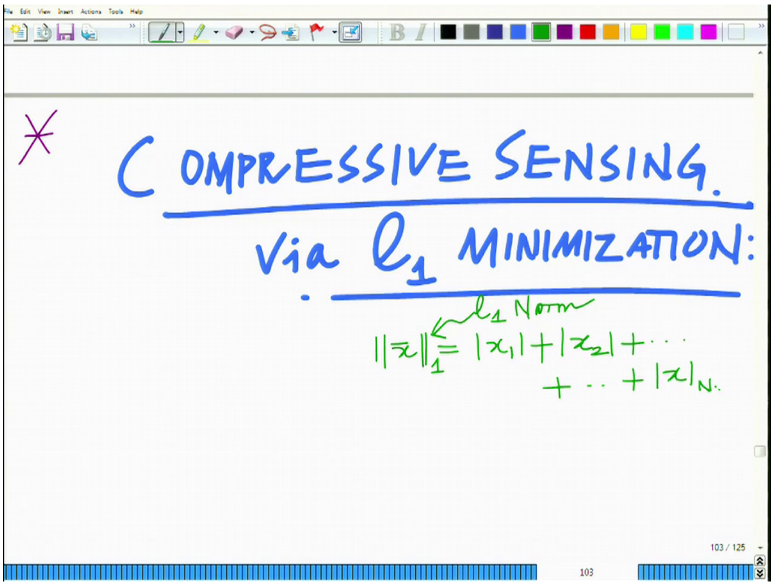
Applied Optimization for Wireless, Machine Learning, Big Data
Prof. Aditya K. Jagannatham
Department of Electrical Engineering
Indian Institute of Technology, Kanpur

Lecture – 59
Practical Application: L1 norm minimization and Regularization approach for Compressive Sensing Optimization Problem

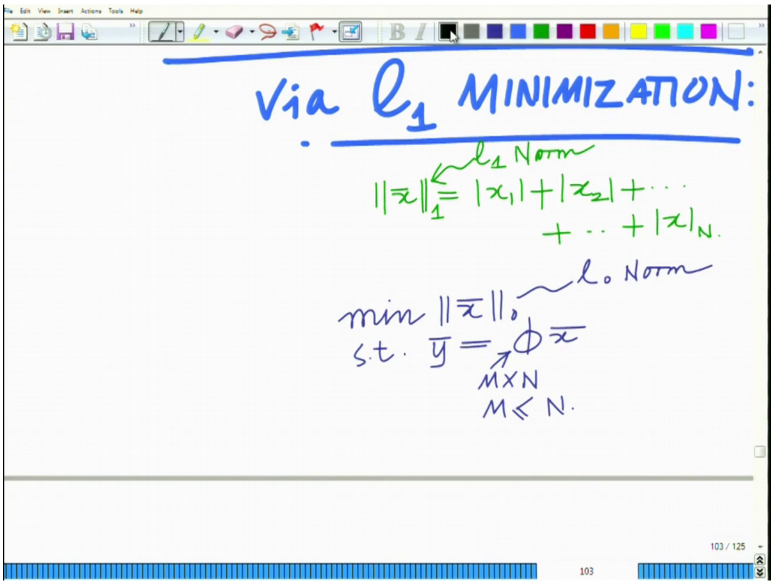
Hello, welcome to another module in this massive open online course. We were looking at compressive sensing, and we have discuss the orthogonal matching pursuit, we also seen an example for the same. Let us look at another totally, completely different and radical approach to for this to tackle this compressive sensing problem, and that is as follows.

(Refer Slide Time: 00:34)



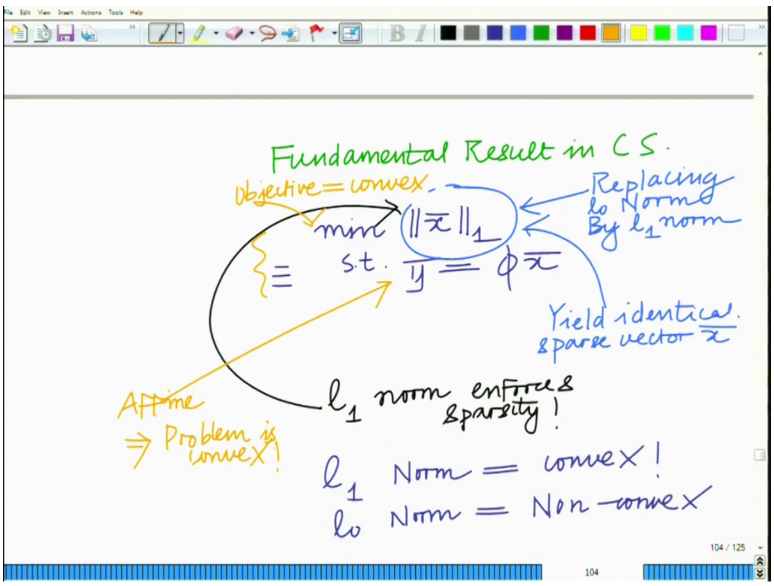
So, we want to look at compressive sensing via l 1 that is you want to minimize the l 1 norm l 1 minimization. And remember, we have seen what is the l 1 norm, the l 1 norm of a vector that is if you have an n-dimensional vector x bar, the l 1 norm the l 1 norm is simply the sum of the magnitudes of the components of x bar, this is the l 1 norm ok. And so this is the l 1 norm of the vector x bar, this is sum of the magnitude is the components of x bar.

(Refer Slide Time: 01:56)



Now, we have seen the compressive sensing problem is the following thing that is what we are doing is we are minimizing norm l 0 norm subject to the constraint that the observations or measurements y bar equals phi times x bar. Phi is the sensing matrix we have seen, which is M cross N number of observation. Number of observations M is less than or equal to the number of unknowns N. And one of the fundamental results in fact this is the most fundamental result in compressive sensing, it states interestingly that this l 0 norm minimization can be replaced by l 1 norm minimization, and still you can recover the sparse vector x bar ok.

(Refer Slide Time: 02:54)



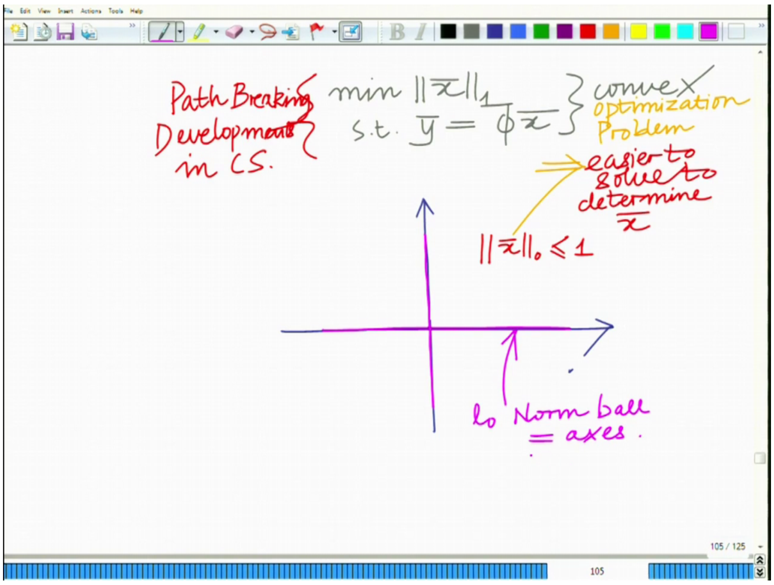
So, the fundamental result in compressive sensing that is instead of minimizing the l 0 norm, you can replace that so the above problem is equivalent to replacing the l 0 norm by replacing the l by the l 1 norm subject to again the same constraint that is so what we are doing is we are replacing l 0 norm by the l 1 norm. And this problem is equivalent to the earlier optimization problem or for a large number of scenarios can be shown to yield with very high probability, yields the identical sparse vector x bar.

So, what this says is the l 1 norm also enforces sparsity. The l 1 norm is also equally good at enforcing sparsity. So, what we can do is we can replace the l 0 norm. Remember l 0 norm, it is very difficult, because it is simply related to the number of non-zero elements in x bar. Instead of that we are replacing it by the l 1 norm. And it can be shown that for a large number of scenarios or with very high probability, the solution x bar that is obtained as a solution of both these above optimization problems is the same.

So, then what is the advantage in replacing the l 0 norm by the l 1 norm. The significant advantage is, if you remember the l 1 norm is convex in nature all right. So, if you look at the l 1 norm, the l 1 norm is convex, the l 0 norm is highly non- convex. So, therefore if you look at this optimization problem over here, the objective is convex, and the constraint is of course affine constraint, which is convex implies the problem is convex. So, what we are achieving by this is that we are converting a problem, which was previously highly non-convex into something that is convex.

So, this is a convex optimization problem. And as we said a convex optimization problem can be solved relatively easily and accurately. And therefore, it is much easier to solve this problem. And determine the sparse vector x, so this is in fact a boon all right, so that these two problems are equivalent, because the previous one is highly non-convex well this is convex. It can be solved relatively easily, and one can find the sparse vector x.

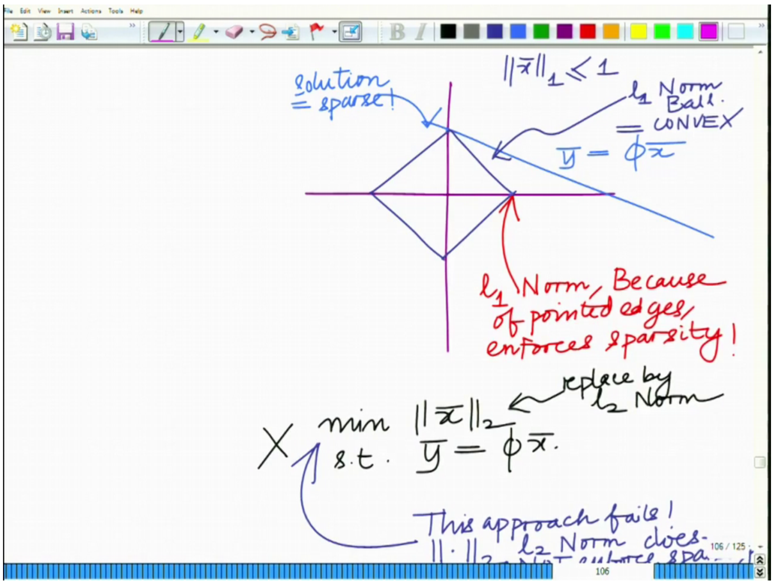
(Refer Slide Time: 06:48)



So, this minimization of norm x 1, which is the CS problem compressive sensing problem subject to y bar equals phi times x bar. This is a convex optimization problem implies it is easier to solve to determine x bar. And therefore, this is the basically this is something that is a radically different approach than minimizing the l 0 norms; it is in fact what I say it is one of the most path breaking developments. This is one of the path breaking developments in compressive sensing that is demonstrating that the l 0 norm minimization is equivalent in a large number of scenarios to the l 1 norm minimization.

And you can see this as follows. For instance, if you look at the l 0, so remember we looked at the l 0 ball l 0 norm ball that is for instance, if you consider norm x l 0 norm less than or equal to less than or equal to 1, the l 0 norm ball is simply the axis the l 0 norm ball is simply along the axis. So, this is highly non-convex.

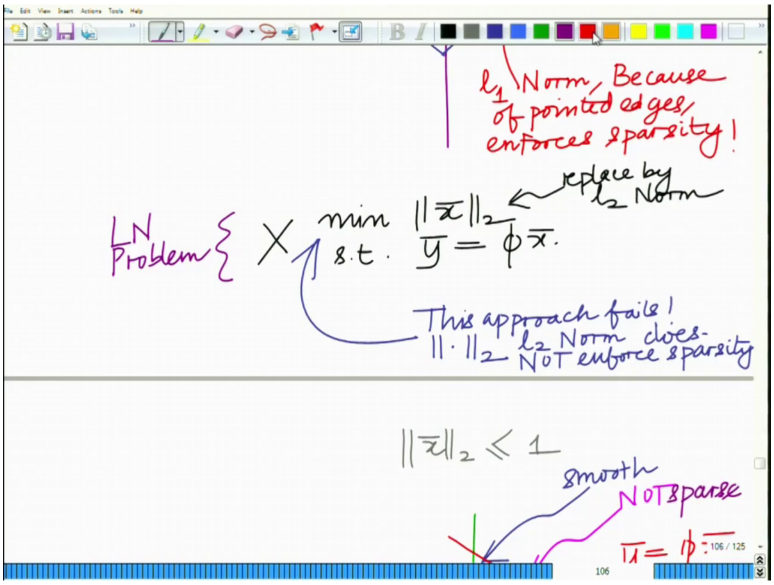
(Refer Slide Time: 09:05)



However, if you look at the l 1 norm ball, the interesting idea about this is that remember norm x bar 1 less than or equal to 1, the l 0 it looks like a diamond. This is this diamond shaped object. This is the l 1 norm ball, which is convex ok; you can see this is a convex shape. Because, if you look at any two points inside, inside this object join them the line segment entire line segment lies inside.

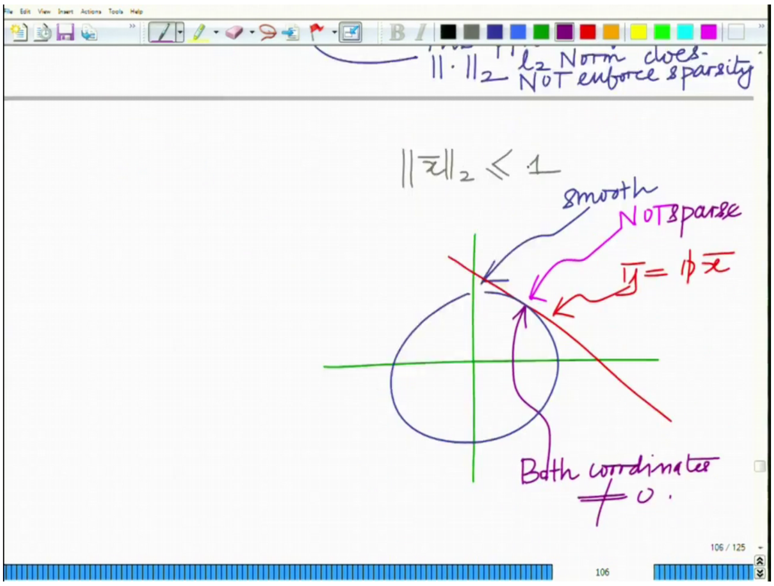
And now if you enforce this constraint a fine constraint, which is nothing but a line y bar equals phi x bar, it intersects at one of these pointed edges that is it lies on the axis implies the solution is sparse ok, because of the axis on the y axis x component is 1, x axis y component is 0 all right. And this you can imagine this in n-dimensional. When it intersects on this one of these pointed edges, several of these components are going to be 0. Therefore, it enforces sparsity similar to the l 0 norm. So, the l 1 norm, because of these pointed edges the l 1 norm because of the pointed edges enforces sparsity.

(Refer Slide Time: 10:56)



Now, unfortunately if you do replace it by the l 2 norm, which is even more convenient, because remember this is nothing but the minimum norm that is replaced by l 0 norm l 2 norm, this is not equivalent. In fact, this approach interestingly this approach fails, because the l 2 norm does not l 2 norm does not enforce sparsity l 2 norm does not enforce sparsity.

(Refer Slide Time: 12:12)



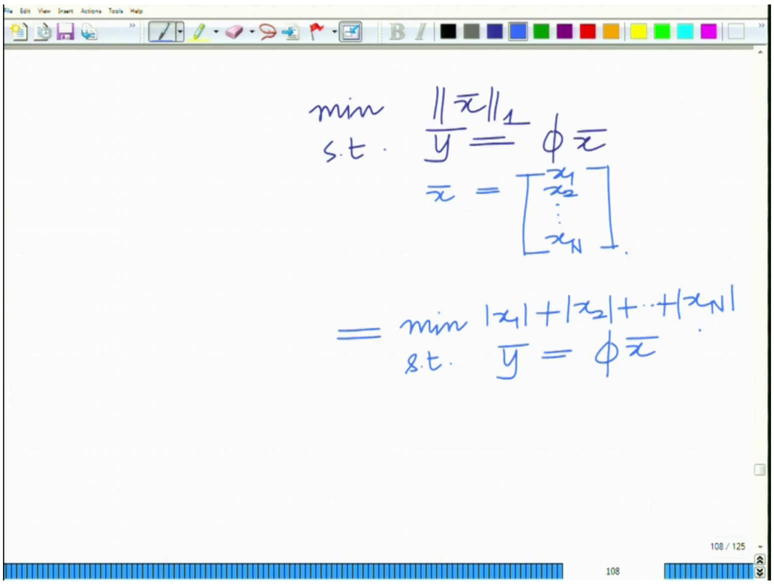
In fact, if you look at the l 2 norm ball if you look at the l 2 norm ball, it looks something like this. It has smooth no pointed edges, it is smooth. And therefore, if you look at, this line y equal to x bar that is the affine constraint y bar equal to phi x bar. This intersects at a point not sparse. The solution you can see, it has both non-zero x and y coordinates. Previously if you look at the l 1 norm ball, it intersects on the axis. When you intersect on the axis, either x coordinate or y coordinate is 0, which means the resulting solution is sparse.

But, if you look at this, because the edges are not pointed, edges are smooth. It will intersects not on the axis, but intersects it is a tangent somewhere in between all right. And therefore, it has both non-zero x and y coordinates, which means the resulting solution is not sparse. So, the l 2 norm does not (Refer Time: 13:18). Of course, there is a simple schematic or let us say a diagrammatic illustration or it more intuitive explanation what is happening. The real proof rigorous proof is also given, but it is beyond the scope of this course all right.

So, this intersects at one of the points in between, so this is not sparse. You can see, because both coordinates are not equal to this is so. This cannot be represented by the although, we would like to represent replace it by the l 2 norm all right, because for the l 2 norm it is much more easier to solve all right, in fact for this problem. This is nothing but the least norm problem right. So, this is basically if you remember this is the least norm problem l n, you have the least squares and you have the least norm. There is a least two norm solutions subject to the constraint y bar equal to phi, you have a closed form solution for this.

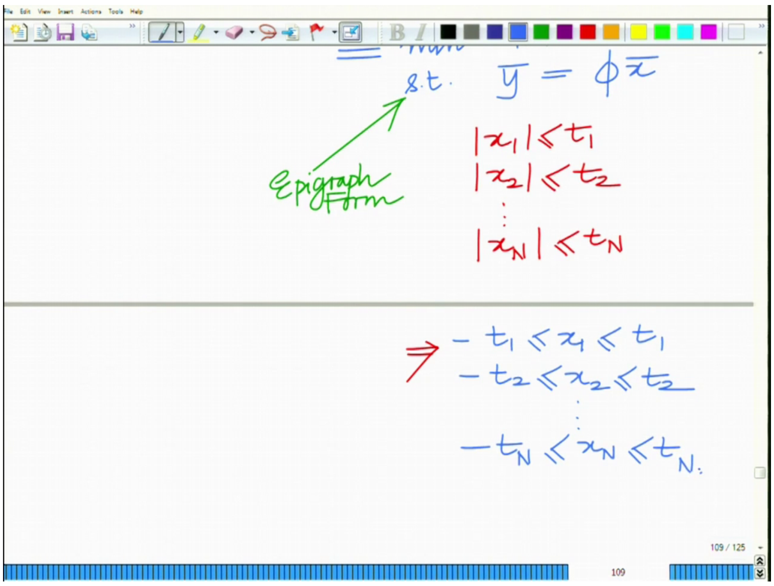
However, this approach fails to recover the sparse signal vector x bar, because the l 2 norm does not only the l 1 norm enforces sparse city. And this is in fact, the revolutionary development of compressive sensing that is this l 0 norm minimization, which was highly non-convex all right. And highly complex problem to solve can be replaced by this convex optimization problem, which is extremely easy, which is relatively easy to solve. And recover the sparse vector x bar, which is identical to that of the previous l 0 norm optimization problem for a with very high probability, and that is in fact the revolutionary development of compressive sensing.

(Refer Slide Time: 15:07)



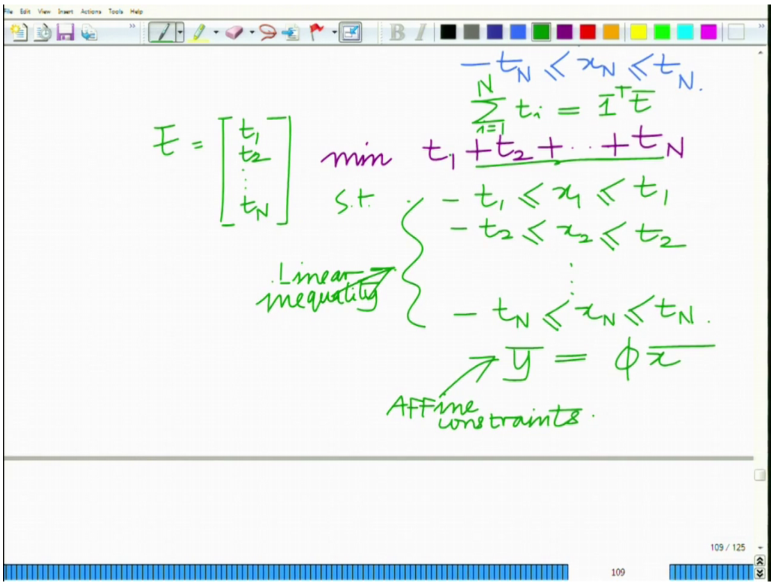
And in fact if you look at the l 1 norm minimization problem one can further simplify it as follows that is you can use the epigraph form, remember you have minimize norm x bar 1 subject to the constraint y bar equal to phi x bar. Now, let us say x bar is an n-dimensional vector, then this problem is basically nothing but minimize the l 1 norm, which is magnitude x 1 plus magnitude x 2 plus magnitude x n subject to the constraint y bar equal to phi x bar.

(Refer Slide Time: 16:02)



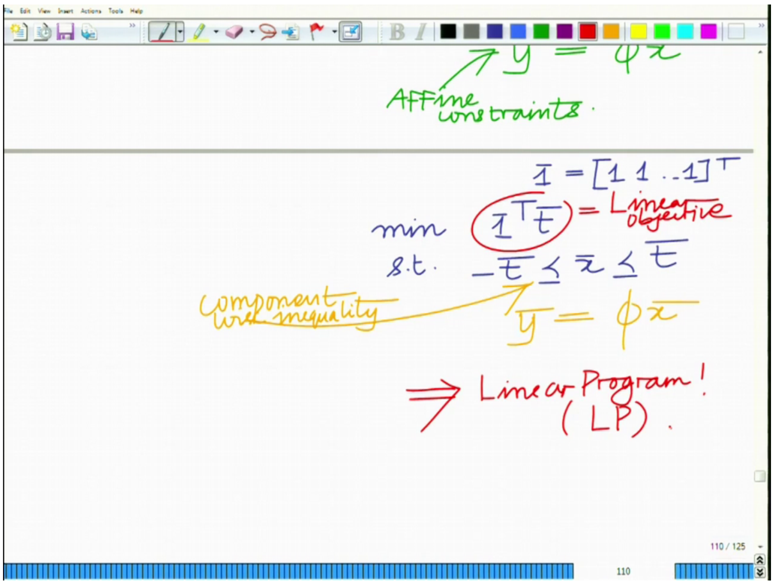
And now, we can use the epigraph form to simplify this. And that is given as follows, you can write magnitude x 1 less than or equal to t 1, magnitude x 2 less than or equal to t 2 so on and so forth, magnitude x N less than or equal to t N, which implies minus t 1 less than or equal to x 1 less than or equal to t 1, minus t 2 less than or equal to x 2 less than or equal to t 2 so on and so forth, minus t N less than or equal to x N less than or equal to t N.

(Refer Slide Time: 17:01)



And therefore, you can write this optimization problem minimize norm x 1 bar as minimize t 1 plus t 2, remember this is epigraph form t 1 plus t 2 plus up to t N. Subject to the constraint your set of linear inequalities minus t 1, this is a known as also box constraint minus t 1 less than or that is x 1 lies between minus t 1 and t 1. Similarly, x 2 lies between minus t 2 and t 2 and so on and so forth, x N lies between minus t N and t N. And you have the original constraint is there ok. So, these are your inequality or linear inequality constraints, and this is your affine. And you can see this is nothing but sum of t bar summation of t i, i equals 1 to i equals 1 to N. And you can also write it as 1 bar transpose t bar, where t bar is your vector.

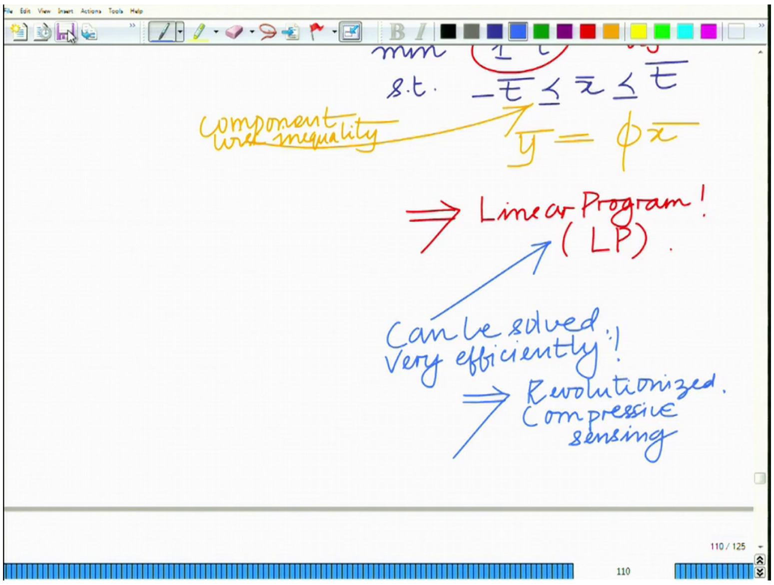
(Refer Slide Time: 18:42)



In fact, you can also write this as minimize 1 bar transpose t bar, where 1 bar is the vector of all 1s, simply performing the sum of the elements of t bar. Subject to the constraint minus t bar less than or equal to x bar less than or equal to t bar, remember it is a component wise inequality ok. We have already seen this squiggly less than or equal to so this is your component wise inequality, just summarizing compactly representing the earlier. So, this is each component of minus t bar less than equal to each component of x bar, which is in turn less than or equal to each component of t bar. So, this is your component wise inequality.

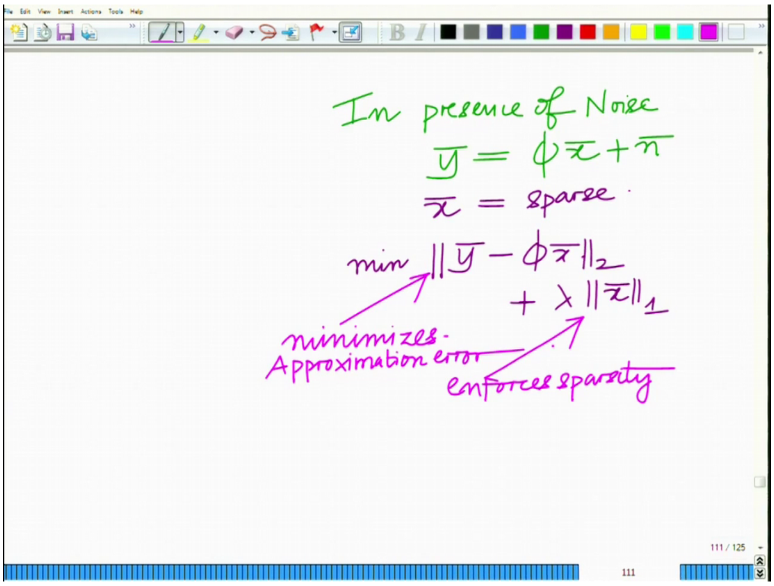
And of course, the linear constraint remains y bar equals the affine constraint. And you can now see these are linear inequalities affine constraint linear objective, objective is linear. So, implies this is a linear program. So, the compressive sensing problem to estimate the sparse vector x bar reduces to a linear program, which for which there are efficient techniques to solve all right.

(Refer Slide Time: 20:30)



From the very complex l 0 norm optimization, you have reduced it to l 0, l 1 norm optimization, and then in turn to a equivalent linear program. And this is can be solved very efficiently this can be solved very efficiently, so this can be solved very efficiently. And this is what has revolutionized, this is what as revolutionize compressive sensing ok. So, it is conversely ok.

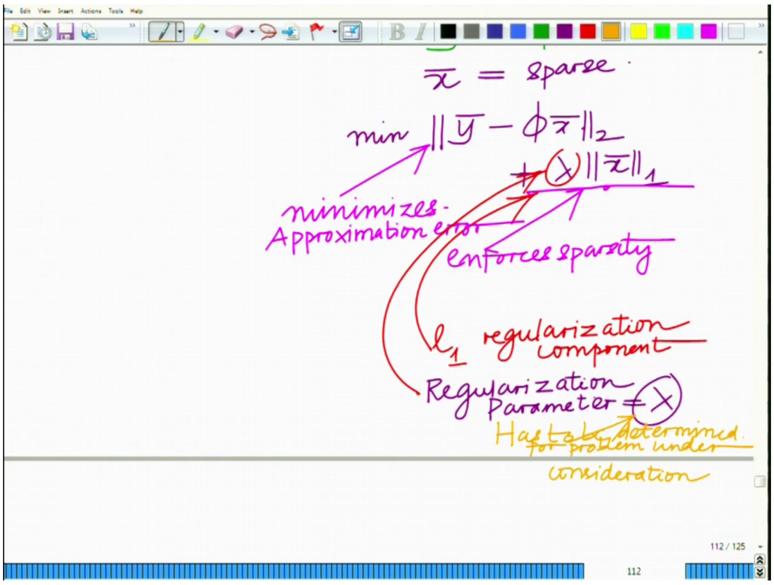
(Refer Slide Time: 21:18)



Now, in the presence of noise, now what happens? Now so far we are considered a noiseless observation model. Now, what happens to it in the presence of noise, it is very simple. So, you have the observation model y bar equals phi x plus n bar, you can no longer write y equals y bar equals phi times x bar. And the vector x bar equals sparse. Now, no previously when you have y bar equal to phi times x bar, remember you minimize the least squares that is you minimize previously, you minimize y bar minus phi x bar the two norm y bar minus norm of y bar minus 1.

Now, in addition you can minimize plus lambda times norm x bar 1, because x bar is 1. And we have seen that this enforces sparse city this enforces sparsity. So, you have the two norm minimization all right to. So, this minimizes the model error or minimizes the approximation or fit error or the observation model error and this enforces sparsity.

(Refer Slide Time: 22:59)



So, basically you are adding an l 1 regularization component. This is termed as an l 1 regularization component; we are adding an l 1 regularization component. And this lambda is termed as the regularization parameter this lambda is a regularization parameter. And this has to be determined for the problem under consideration this has to be determined for the problem under consideration.

So, you are taking the classic least squares that is minimizing the l 2 norm of the approximation error plus a regularization term all right that is lambda is the regularization parameter times the norm 1 of x bar, which enforces sparsity of the vector x. So, recovering the sparse vector as well as at the same time minimizing the approximation error all right.

So, basically this is known as the regularized. So, you can talk about as l 1 regularized. So, you can this is regular this is your regularization component. And so you can turn this as basically your regularized version of the previous problem all right, where you have a noisy observation model, and you simultaneously want to enforce sparsity to estimate a sparse signal vector x all right.

So, basically that sort of summarizes our discussion, it is a very high level discussion. And this the theory of compressive sensing is very deep, where you demonstrate that the l 0 norm either you demonstrate the l 0 norm optimization is equal to or l 1 norm optimization is equal to l 0 norm optimization or the various applications of compressive sensing or in might numerous, there are several and several applications are still coming out. And they continued come out for compressive sensing, but this has been a significant and path breaking recent development in the field of signal cross all right.

So, I hope it takes some time to explore the various applications of compressive sensing for different paradigms, be it communication, be it signal processing, be it detection or beat it signs, so several applications all right. It has several applications and problems, which is previously unsolved or previously not solved very efficiently can now using the framework of compressive sensing have been shown to be solved very efficiently with much better performance all right. We will stop here.

Thank you very much.